SCREENING FOR THE FUNCTIONALITY OF RNA TEMPLATED REPAIR OF DOUBLE-STRAND DNA BREAKS IN *SACCHAROMYCES CEREVISAE*

A Thesis
Presented to
The Academic Faculty

by

Katherine Gordon

In Partial Fulfillment
of the Requirements for the Degree
B.S. in Biology with Research Option in the
School of Biological Sciences

Georgia Institute of Technology
May 2017

SCREENING FOR THE FUNCTIONALITY OF RNA TEMPLATED REPAIR OF DOUBLE-STRAND DNA BREAKS IN *SACCHAROMYCES CEREVISAE*

Approved by:

Dr. Francesca Storici, Advisor
School of Biological Sciences
Georgia Institute of Technology

Dr. Mirjana Brockett
School of Biological Sciences
Georgia Institute of Technology

Date Approved: August 1, 2016

ACKNOWLEDGEMENTS

I would like to thank the School of Biological Sciences, and in particular Dr. Storici and her lab. Specifically, I would like to thank Chance Meers for mentoring me and assisting me with this project. I would also like to thank Dr. Brockett for being my second reader, and for supporting me throughout the project.

TABLE OF CONTENTS

LIST OF FIGURES AND TABLES

LIST OF SYMBOLS AND ABBREVIATIONS

DSB	Double-Strand Break
S. cerevisiae	*Saccharomyces cerevisiae*
E. coli	*Escherichia coli*
His^-	Medium Containing no Histidine
Ura^-	Medium Containing no Uracil
YPGal	Medium Containing Yeast Extract, Peptone, and Galactose
YPD	Medium Containing Yeast Extract, Peptone, and Dextrose

SUMMARY

Double-strand breaks (DSBs) in DNA are detrimental, as they can cause mutations and genomic rearrangements, which in turn leads to cancer and other diseases[1]. Recent research reveals that DSBs can be repaired by RNA-templated homologous recombination[2,3]. This finding expands the known instances in which RNA can be converted into DNA. However, RNA-templated repair of DSBs is not well understood. In order to better understand the mechanism of RNA-templated repair of DSBs, the current research aims to identify the proteins that facilitate the repair. The research utilizes a system wherein RNA-templated repair of DSBs is known to occur. A yeast overexpression plasmid library was produced in order to test the ability of fragments of the yeast genome to facilitate RNA-templated repair of a DSB when these are highly expressed in the yeast cells. In order to test the ability of added gene fragments to facilitate RNA-templated repair of DSBs, the experimental candidates were exposed to galactose in order to induce a DSB, to activate the transcription of the overexpressed gene fragment, and to initiate the transcription of anti-sense RNA used to repair the break. The candidates were then moved to medium without histidine in order to assess the frequency of repair. Once a large number of colonies (~50,000) are screened, we expect to identify several proteins that facilitate RNA-templated repair of DSBs. Identifying the specific genes that facilitate this repair mechanism will assist in characterizing the functionality of the RNA-templated DNA repair mechanism. Identifying these genes will also allow for better predictions for how this same phenomenon may occur in human cells.

CHAPTER 1

INTRODUCTION

When double-strand DNA breaks (DSBs) occur, cells can activate non-homologous end joining which adds two pieces of DNA containing a DSB together. This process can lead to mutations or genomic rearrangements that can cause cancer and other diseases[1]. However, DSBs can also be repaired by homologous recombination, which mitigates these risks. Because of this, it is important to understand what mechanisms are used by cells to repair the breaks. Currently, there is knowledge of several mechanisms by which DSBs can be repaired. Specifically, researchers are studying the repair of DSBs through mechanisms that involve RNA. One process of DSB repair involves microRNAs, which are small RNAs that can target genes involved with the checkpoints that a cell passes through before division[4]. Another way RNA can be involved in DSB repair is through the recruitment of DNA repair factors. This mechanism is initiated by a different kind of small RNA, called DSB-induced small RNAs (diRNAs), which are produced near sites of DSBs[5]. In addition to the mechanisms of DSB repair that involve the use of small RNAs as regulatory tools or as recruiting factors, there is also evidence that RNA can be used as a template for DSB repair via homologous recombination[2,3].

The ability for RNA to act as a template during DSB repair challenges the central dogma of molecular biology. Originally, it was thought that RNA is only converted to DNA during limited situations such as telomere synthesis, [6] and retroelements, [7] but the evidence that RNA can act as a template for homologous recombination indicates that this conversion is more widespread. Thus, it is important to further explore the conditions that facilitate the conversion of RNA to DNA during the RNA templated repair of DSBs. Identifying the proteins involved with the functionality of this mechanism would not only establish more validity of the mechanism in the field, but it

would also reveal the specific conditions in which RNA can be converted to DNA and used for repair within a eukaryotic cell. Such advancements could lead to developments in gene therapy in order to cure and prevent genetic diseases.

CHAPTER 2

MATERIALS AND METHODS

Yeast Strain

In order to induce RNA templated repair of DSBs in DNA, a specific yeast strain was used. This particular yeast strain contained a molecular system developed by Keskin, *et al.*,[2] which is known to exhibit RNA-templated repair of DSBs. The yeast strain used, called YS486, contained an artificial intron, which was oriented in the antisense direction. This artificial intron was integrated within the non-functional *his3* gene on chromosome III of *S. cerevisiae.* The system also included a homothallic switching endonuclease in order to induce a DSB within the artificial intron. Since only splicing from the *his3* antisense RNA results in the removal of the artificial intron, the galactose promoter was included which enabled the cells to transcribe antisense RNA. Since RNase H1 and RNase H2 degrade RNA within RNA-DNA hybrids, the yeast strain contained markers in place of the genes associated with RNase H1 (*RNH1)* and the catalytic subunit of RNase H2 (*RNH201).* This prevented antisense RNA paired with DNA from being degraded. Since Spt3 is involved in the reverse transcription of RNA to cDNA, the experimental yeast strain used included a marker in place of the gene responsible for the production of Spt3 (*SPT3).*

Overexpression Screening

In order to identify which proteins facilitate the RNA templated repair of DSBs, a yeast overexpression plasmid library was utilized. Each plasmid of the library contained a fragment of the yeast genome. In order to initiate the transcription of this fragment, the plasmid also contained the galactose promoter. A Ura3 marker was included to confirm that the tested candidates contained the plasmid. This overexpression plasmid library was developed by the Ron Davis lab of the Stanford University School of Medicine via

the protocol explained in their 1990 paper[8]. These plasmids were then extracted from *Escherichia coli* using a GeneJET plasmid Midiprep kit. The plasmids were then transformed into YS486 via the lithium acetate protocol.

Candidates Tested

In order to screen individual candidates, colonies, each containing a different plasmid from the library, were placed in every well of a 96 well plate. In order to validate the candidates as viable as well as to compare the overexpressed candidates with a baseline, four controls were used. As a positive control, a His^+ wild type colony containing YCp50pk was used. In order to determine which candidates contain gene fragments that facilitate RNA-templated repair of DSBs, YS486 containing a YCp50pk plasmid with just the Ura3 marker was used. A *spt3* yeast strain was used as a negative control, since cells without *SPT3* cannot convert RNA into cDNA. Finally, as a positive control, *rnh1 rnh201 spt3* (His^+) + pAC3013 was used. Plasmid pAC3013 contains a yeast gene that, when overexpressed, imitates the frequency of DSB repair by RNA.

Selective Media

In order to assess the ability of candidates to promote repair of DSBs via RNA-templated repair, they were plated on a series of selective media. After placing the candidates in a 96 well plates, 43 candidates and the controls were pronged (Figure 1) onto medium without uracil (Ura^-), and medium without histidine (His^-). After two days of growth, the candidates were replica plated from the Ura^- plate to a plate containing galactose (YPGal). The candidates grew on YPGal in order to induce a DSB, turn on the transcription of antisense RNA, and turn on the transcription of the gene fragment within the plasmid. After two days of growth on YPGal, the candidates were replica plated onto His^-. After at least two days of growth on His^-, the candidates were observed and compared to the reference control to determine which candidates exhibited more repair, which is seen as higher frequency of His^+ colonies on His^- medium. In order to re-evaluate candidates that appeared to exhibit more repair than the reference control, a

patch assay was conducted. These candidates were streaked (Figure 2) from their original Ura$^-$ plate onto a rich medium called YPD to isolate single colonies. After 3 days of growth, single colonies were patched (Figure 3) onto Ura$^-$. Following two days of growth on Ura$^-$, these candidates were replica plated to YPGal and then after two days of growth, replica plated to His$^-$ for secondary comparison to the reference control.

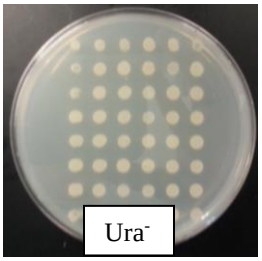

Figure 1: All tested candidates and five control strains were added to a 96 well plate. 48 candidates were pronged to a Petri dish containing medium without uracil.

Figure 2: Candidates that exhibited more growth than a control containing an empty plasmid were streaked onto a Petri dish containing a rich growth medium called YPD in order to isolate single colonies.

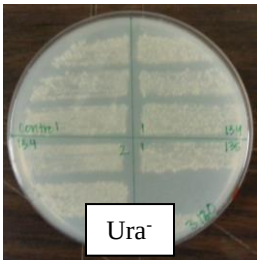

Figure 3: Three single colonies from each successful candidate were obtained from YPD plates, and patched in rectangles on a Petri dish containing medium without uracil.

DNA Extraction and Sequencing

Candidates that consistently appeared to exhibit more repair than the reference control were identified, and sequenced. In order to determine the identity of the gene fragment that was included in the plasmid of successful candidates, the DNA of the plasmid was extracted and amplified. They first grew in Ura$^-$ liquid medium and were extracted. Transformation into *E. coli* was conducted to isolate plasmids, which were then extracted. Primers (Figure 4) were added to these plasmids before they were sequenced. The DNA was sequenced by Eurofin Genomics using the Sanger method[9].

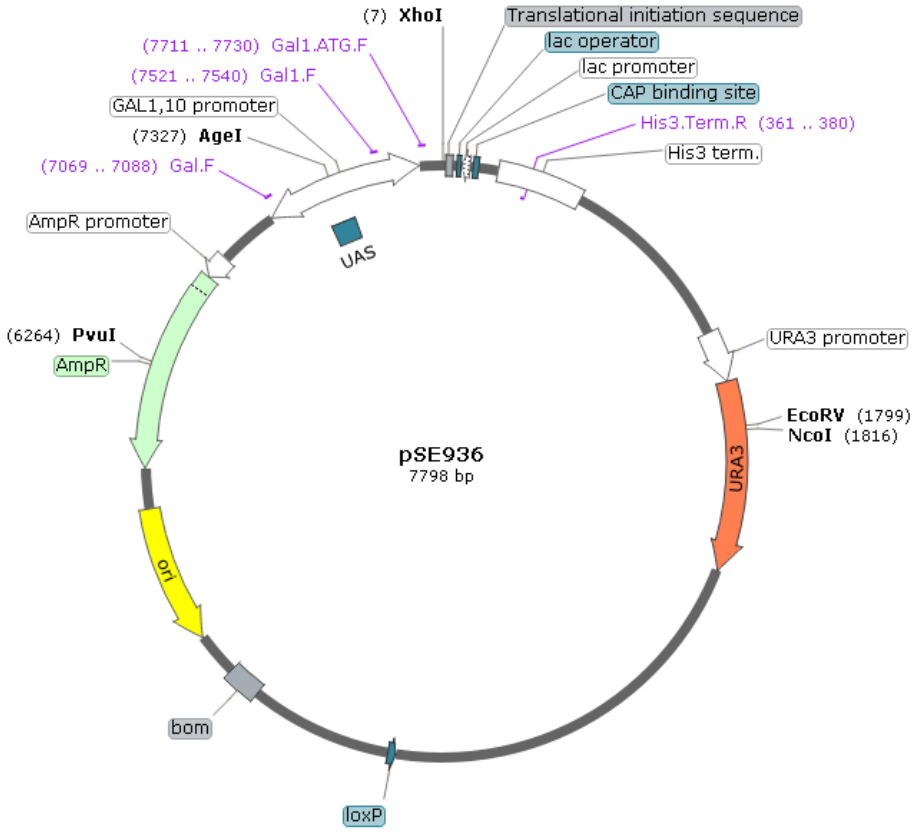

Figure 4: A variety of primers were used to sequence the plasmids included in the overexpression plasmid. Primers are indicated by purple font.

CHAPTER 3

RESULTS

Successful Candidates

After screening around 15,000 candidates, about 550 indicated more repair after a first trial of selective media, as compared to a control containing an empty plasmid. However, after two additional rounds on selective media, six of these indicated more growth than a control containing an empty plasmid. Figure 5 shows these successful candidates, and displays the differences in amount of growth between candidates, as well as compared to a control containing an empty plasmid.

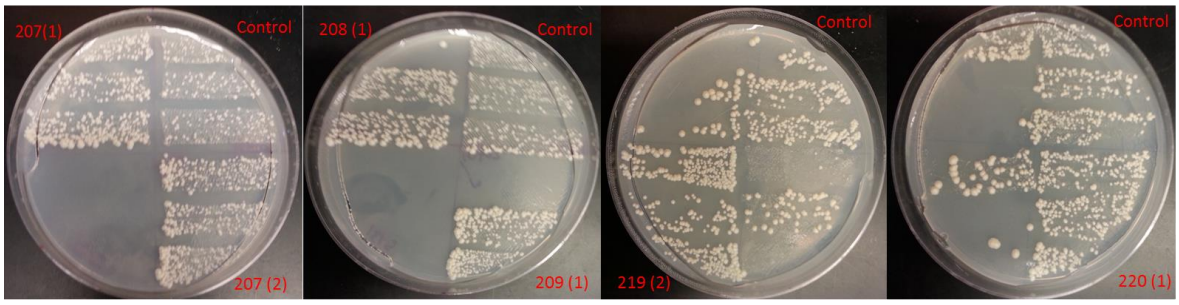

Figure 5: Six candidates exhibited more repair than a control containing an empty plasmid. After candidates were selected from Ura$^-$ they were streaked onto YPD to isolate single colonies, then patched onto Ura$^-$. They were then replica plated to YPGal and then His$^-$ in order to assess repair. The control used was YS486+Ycp50pK. Candidates shown are 207(1), 207(2), 208(1), 209(1), 219(2), and 220 (1).

Sequences of Successful Candidates

The sequence results for the six tested candidates are displayed in Figure 6. The sequences shown were determined by the His primer. While additional primers were also sequenced, they were not successful. Since the sequences were determined using only one primer, the sequences displayed in the figures show the gene sequences from only

one direction. As seen in Figure 6, every candidate tested contained a different sequence within the overexpression plasmid. The size of these fragments varied among samples, as did the identity and quantity of genes present.

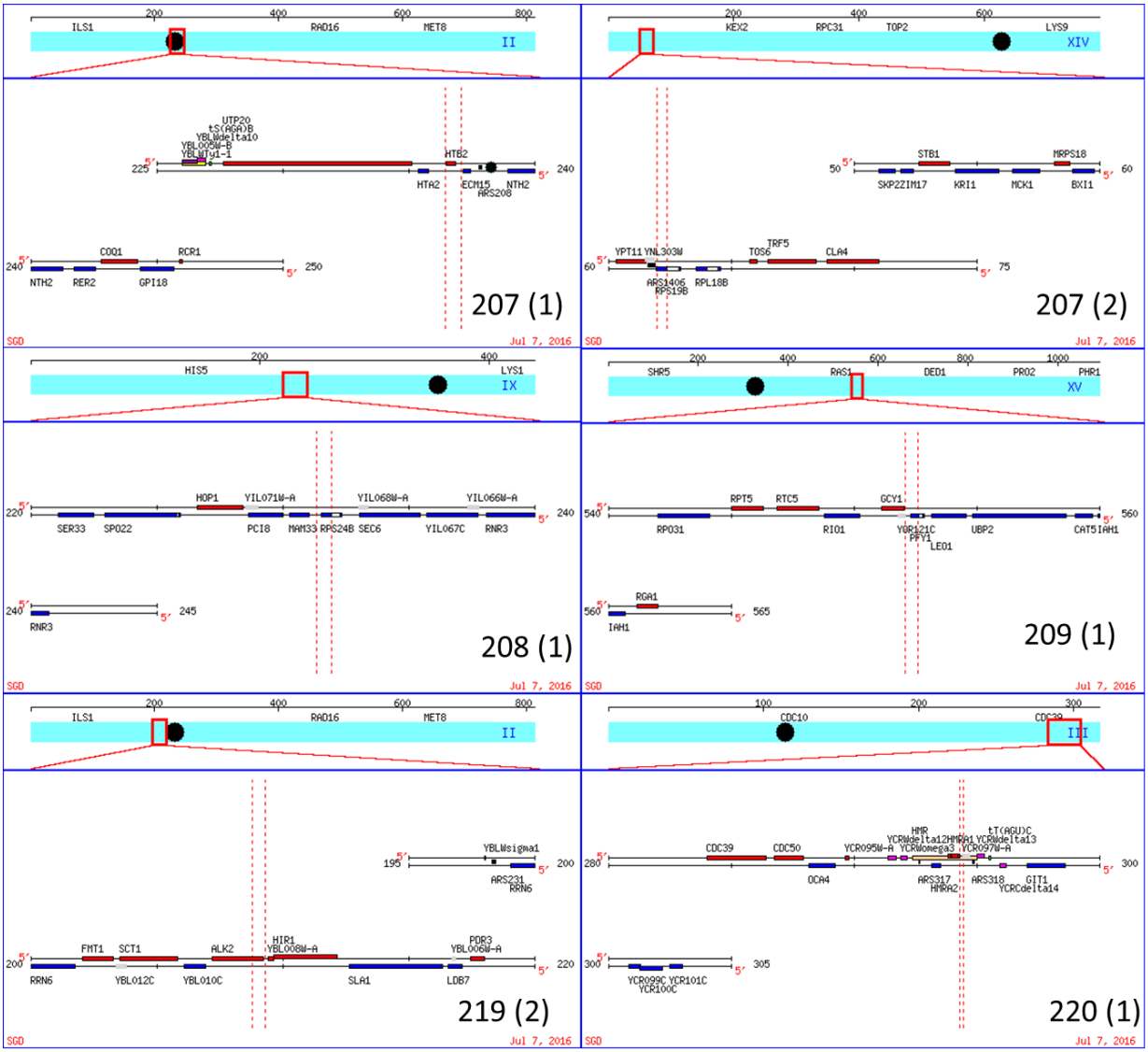

Figure 6: The sequence of the added gene fragments of six candidates were identified using the His primer. Plasmids from successful candidates were extracted, transformed into *E. coli*. Plasmids were extracted from *E. coli* and then primers were added to the plasmid DNA for sequencing. The samples were sequenced by Eurofin Genomics. Several primers were tested, but only His produced reliable sequencing data.

Results shown reveal the sequences of genes contained within the plasmids for one, but not both directions.

Identified Gene Fragments

From the six successful candidates showing more repair, several coding genes were identified from the overexpression plasmids. From these six candidates, five were sequenced successfully by the His primer. The genes identified and their known function are displayed in Table 1. While the sequences indicate that at least a portion of each of these genes are present in the tested candidates, the sequence does not reveal whether the entire gene or just a gene fragment is present. A variety of gene functions are associated with the identified genes. These functions range from unknown proteins, repeating regions, protein kinases, genes involved in histone synthesis, and even proteins phosphorylated in response to DNA damage.

Candidate	Gene Name	Function
207(1)	HTB2	Histone
207(1)	HTA2	DNA damage-dependent phosphorylation by Mec1p; facilitates DNA repair
207(1)	ECM15	Unknown function
207(2)	ARS1406	Autonomously Replicating Sequence
207(2)	RPS19B	Protein component of the small (40S) ribosomal subunit
207(2)	YNL303W	Dubious open reading frame
208(1)	RPS24B	Protein component of the small (40S) ribosomal subunit
208(1)	MAM33	Specific translational activator for the mitochondrial COX1

		mRNA
209(1)	PFY1	Profilin; protein abundance increases in response to DNA replication stress
219(2)	ALK2	Protein kinase; phosphorylated in response to DNA damage
219(2)	HIR1	Subunit of the HIR complex; involved in regulation of histone gene transcription
219(2)	YBL008W-A	Protein of unknown function

Table 1: The sequence from the overexpression plasmid of 6 successful candidates reveals a variety of proteins that may facilitate RNA-templated repair of DSBs. Plasmids sequenced using the His primer were successful in reporting a viable sequencing data. Sequences of overexpression plasmids from successful candidates were interpreted using the database on yeastgenome.org. A variety of genes were found in each candidate, each with a different known or unknown function. Functions shown in yellow indicate possible facilitative proteins. Results from candidate 220 were omitted due to errors in the sequencing process.

CHAPTER 4

DISCUSSION

Scope of Candidates Found

As indicated in the results section, despite the large amount of candidates tested, only a few consistently exhibited positive results. The *S. cerevisiae* genome is known to be about 12 million base pairs long, and consists of only 3% protein coding genes[10]. As such, it is likely that the current study has only surveyed a small portion of the entire yeast genome, and an even smaller portion of the protein coding regions of the genome. Due to the nature of the overexpression plasmid used, the overexpressed genome fragments were likely to contain only fragments genes which were unable to synthesize functional proteins, and thus, most of the candidates that were tested were unlikely to exhibit higher repair than a control containing an empty plasmid.

Identification of Genes

Although several genes were identified from the successful candidates, they cannot be confirmed as facilitative since their location within the plasmid is unknown. Since the sequences available were produced from one direction only, there is no confirmation regarding the length of the gene insert within the overexpression plasmid. Therefore, it is possible that some of the genes from the successful candidates were not fully expressed, and thus could not have been producing proteins that facilitate RNA-templated repair of DSBs. Regardless, some of the identified genes produce functions that are promising in regards to assisting RNA-templated repair of DSBs, while others are known to facilitate the production of histidine, and could potentially lead to increased rates of growth, as the system in place occurs within the his3 gene. As seen in Table 1, ALK2 codes for a gene that is phosphorylated in response to DNA damage, so it is possible that this a gene that facilitates RNA-templated repair of DSBs. The gene HTA2 is also phosphorylated in response to DNA damage, and is known to be associated with

DNA repair, so it is possible that this protein could facilitate RNA-templated repair of DSBs. Similarly, the genes of unknown function, such as ECM15 and YBL008W-A also have the potential to facilitate RNA-templated repair of DSBs. Overall, it is possible that any, all, or none of these identified proteins actually facilitate RNA-templated repair of DSBs, but with further research and better understanding of which of these genes were present in full within the overexpression plasmids, then the mechanism could be revealed. Unveiling the proteins that facilitate RNA-templated repair of DSBs, also reveals proteins that enable the conversion of RNA to DNA, and thus the reversal of the central dogma of molecular biology[7].

CHAPTER 5

CONCLUSION AND FUTURE WORK

Conclusion

After screening around 15,000 candidates, six exhibited consistently higher repair than the reference control. From these successful candidates, several proteins were identified in conjunction with higher rates of repair. This indicates that RNA-templated repair of DSBs could be facilitated by a variety of proteins. In order to distinguish the facilitative proteins, however, the plasmids from the successful candidates must be sequenced again using an additional primer. This will enable the genes that were fully expressed to be distinguished from those that were only partially present in the overexpression plasmids.

Future Work

In addition to fully identifying the successful proteins for the facilitation of repair of DSBs, more candidates should be tested. By testing more candidates, more proteins can be identified, and thus the mechanism better understood. In addition to testing additional candidates, the plasmids of the successful candidates from this experiment can be transformed into the YS487 strain and tested again. This would further support that the genes incorporated within these candidates may facilitate RNA-templated repair of DSBs by displaying higher frequencies of repair within a strain different from the one used for this experiment. Beyond the scope of this experiment, more work can be done regarding the mechanism. In addition to better understanding how the mechanism works in yeast cells, the analogous mechanisms can also be studied in mammalian cells.

REFERENCES

1. Lieber, M. R. **The Mechanism of Double-Strand DNA repair by the non-homologous DNA end joining pathway.** Annual Review of Biochemistry 79 (2010) pp, 181-211

2. Keskin, H., Y. Shen, F. Huang, M. Patel, T. Yang *et al.,* **Transcript-RNA-templated DNA recombination and repair** Nature 515 (2014) pp, 436-439

3. Storici, F., Bebenek, K., Kunkel, T. A., Gordenin, D. A., & Resnick, M.A. **RNA-templated DNA repair** Nature 447 (2007) pp, 338-341,

4. Yang, Y.-G., and Qi, Y, **RNA-directed repair of DNA double-strand breaks** DNA Repair 32 (2015) pp, 82-85

5. Gao, M., Wei, W., Li *et al.,* **Ago2 facilitates Rad51 recruitment and DNA double-strand break repair by homologous recombination** Cell Res 24 (5) pp, 532-541

6. Greider, C. W. & Blackburn, E. H. **Identification of a specific telomere terminal transferase activity in Tetrahymena extracts** Cell 43 (1985) pp, 405–413

7. Crick, F. **Central dogma of molecular biology**. Nature 227 (1970) pp, 561-563

8. Elledge, S. J., Mulligan, J. T., Ramer, S. W., Spottswood, M., and Davis, R.W. **ƛYES: A multifunctional cDNA expression vector for the isolation of genes by complementation of yeast and *Escherichia coli* mutations** Proc. Natl. Acad. Sci. USA 88 (1991) pp, 1732-1735

9. Sanger, F., Nicklen, F. & Coulson, A. R. (1977) Proc. Natl. Acad. Sci. USA 74 (1977) pp, 5463-5467

10. Engel, S. R., Dietrich, F.S., Fisk, D. G *et al* **The reference genome sequence of *Saccharomyces cervisiae*: then and now**. G3: Genes/Genomes/Genetics 4.3 (2014) pp, 389-398